

Section 4.1: Scatterplots, Association, Correlation

- * Put your data on the front board: Height and shoe size
- * Boys = blue markers
- * Girls = black markers
- ** Get program and groups on your calculator
- ** Turn in Data Exploration 3 (if you didn't yesterday)
- ** Get out 4.1 notes

Scatterplots: Notes

- Shows the relationship between 2 QUANTITATIVE variables
- Can show categorical variables by COLOR OR SYMBOL
- Individuals are represented by the DOTS on the plot
- Explanatory Variable: (independent variable)
 - On the X axis
 - Explains or causes the change in the Y variable
- Response Variable: (dependent variable)
 - On the Y axis
 - Measures the outcome of an experiment or study

Interpreting Scatterplots:

Overall pattern

outliers?

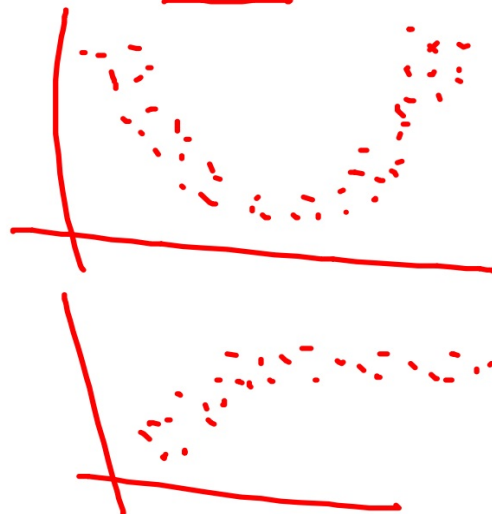
- Get a sense of what the data/plot looks like in general, then ***comment on the following 3 things***

(1) Form

Linear:



Curved:

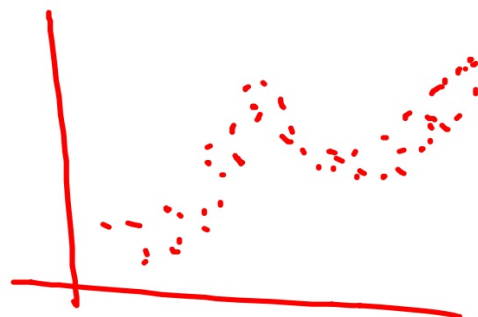


(2) Direction

as $x \uparrow$, $y \downarrow$
Negative Association



as $x \uparrow$, $y \uparrow$
Positive Association



(3) Strength- How well the points fit a form (curved or linear)

- Use the following words (or combinations of these):

- weak
- moderate
- strong
- scattered

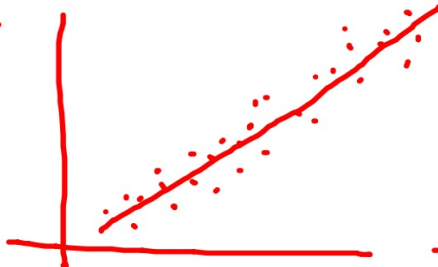


Examples:

Strong



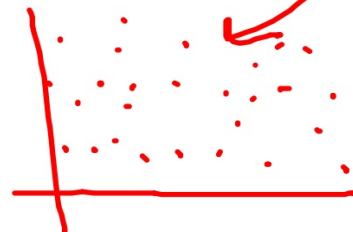
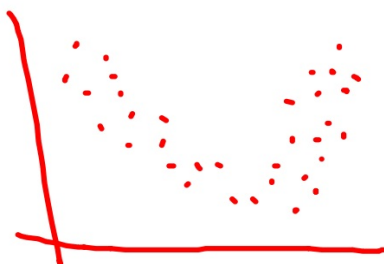
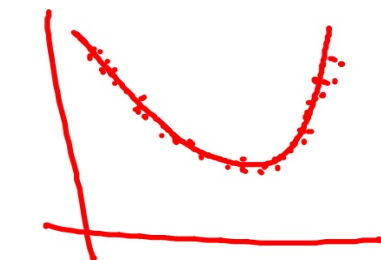
Moderate



Weak

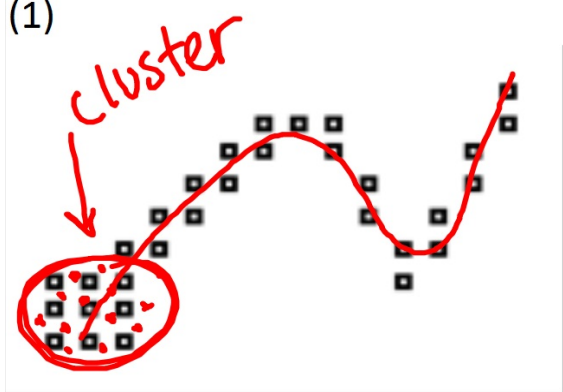


Scattered

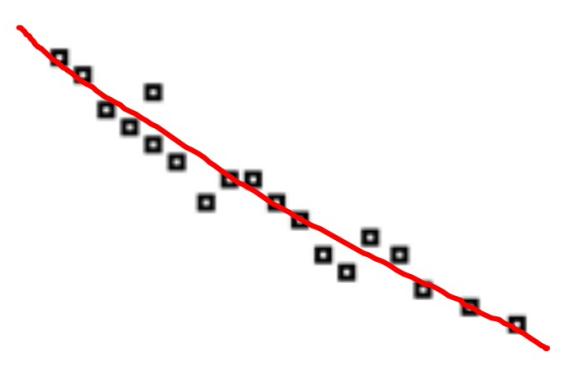


Examples:

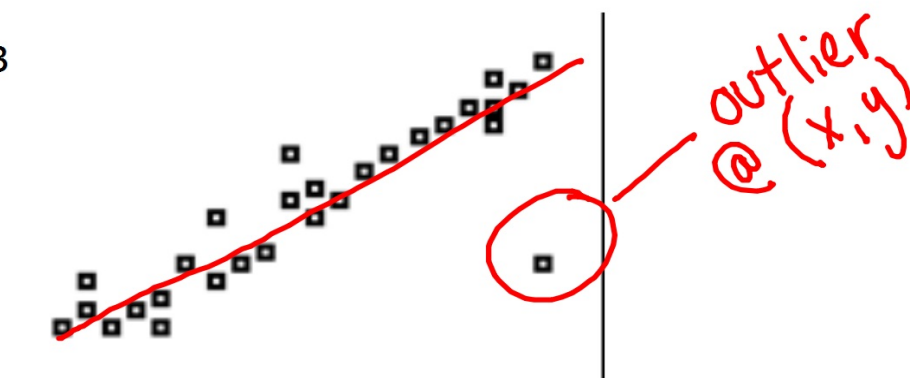
(1)



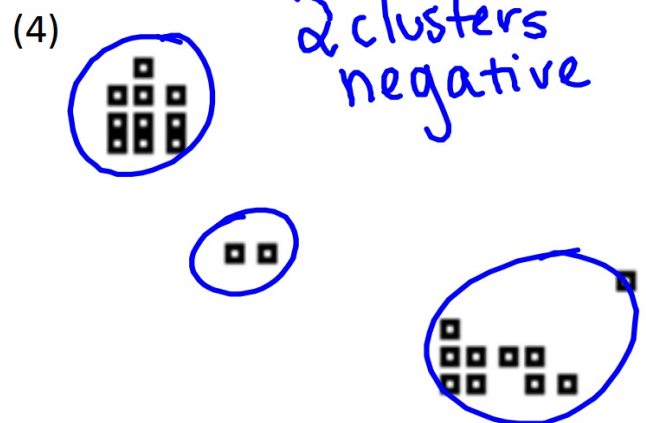
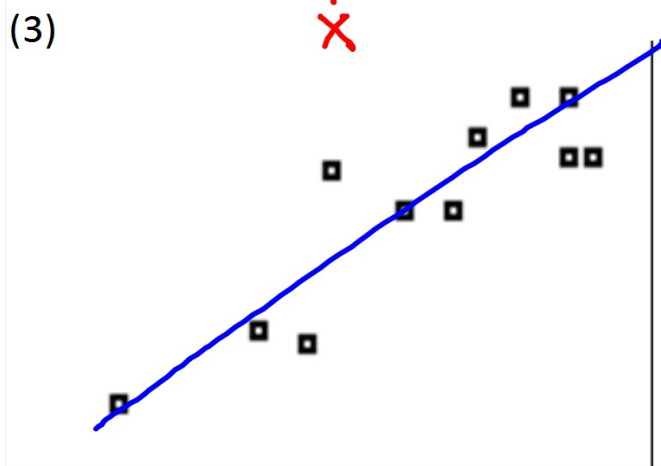
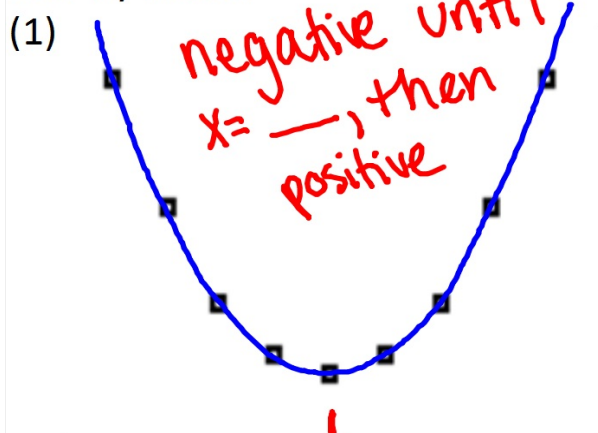
(2)



(3)



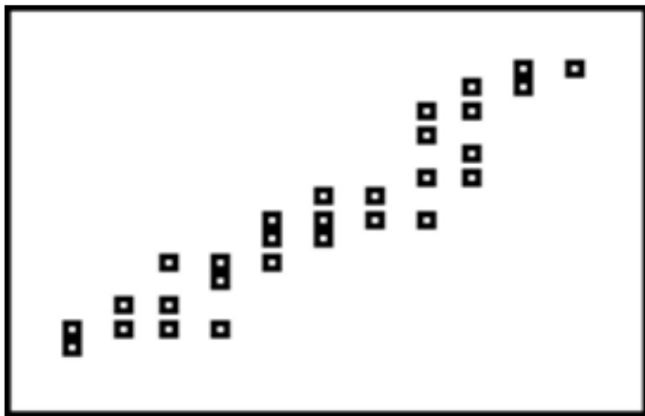
You try these:



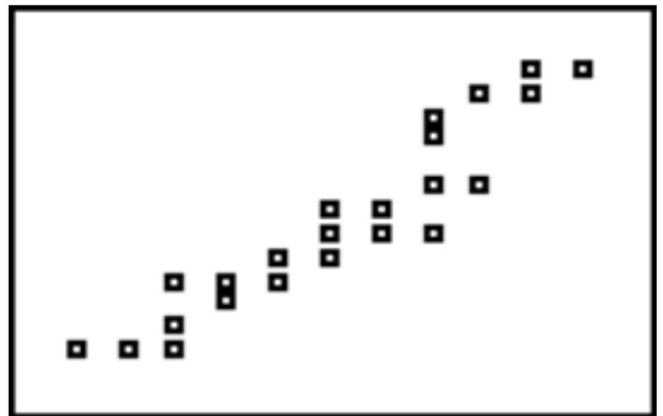
**** Now complete the 4.1A worksheet about describing scatterplots**

WHAT PLOT HAS THE STRONGER RELATIONSHIP? BY HOW MUCH??

A



B



Correlation:

Symbol:

r

Definition:

measures the strength
of the linear relationship
and direction.

~~Formula:~~

$$\frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = r$$

$r = 0.9278$

- * Grouping/Ungrouping
- * How to make scatterplots on the calculator
- * Deleting a point from a list
- * Using the program CORR

**** Have out your HW**

**** Complete worksheet 4.1B (in notes from yesterday)**

HW answers:

1) (a) $x = \text{time}$
 $y = \text{grade}$

(b) $x = \text{height}$
 $y = \text{weight}$

(c) $x = \text{rain}$
 $y = \text{bushels}$

(d) either one can
be x or y

2) (a) $x = \text{income}$
 $y = \text{years of education}$

(b) $x = \text{square footage}$
 $y = \text{price}$

(c) $x = \text{height}$
 $y = \text{arm span}$

(d) $x = \text{amount of snow}$
 $y = \text{amount of water}$

3) (a) positive, linear, weak

(b) (102, 0.2)

(c) A = moderate IQ, low GPA

B = moderate IQ, low GPA

C = low IQ, high GPA

6) (a) and (b) on computer calculator

(c) positive, linear, strong

two seaters had lower numbers overall

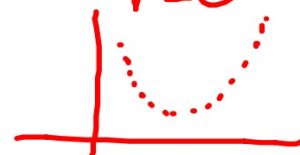
7) (b) positive, linear, strong

8) (a) on computer calculator

(b) weak, negative, linear

Correlation coefficient..... the last few notes.....

- $-1 \leq r \leq 1$
- close to $-1, 1 \Rightarrow$ strong linear
- close to $0 \Rightarrow$ weak linear $r=0$



- r has no units
- affected by outliers
- doesn't matter which Var. is x & y

• changed to x & y

BE CAREFUL!! Correlation does not mean causation!

associated

variable
Don't affect r .

Coefficient of determination:

symbol: $r^2 = (\text{correlation})^2$

$$0 \leq r^2 \leq 1$$

look @ %

* measures % of change in y-variable that's due to change in x-variable.

Ex: $r = 0.82$ $r^2 = 0.6724$



67.24% of change in test scores is due to change in hrs. of sleep.